PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

B1

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G01N 33/53, C12Q 1/68, C07H 21/04

A1

(11) International Publication Number: WO 99/39200

(43) International Publication Date: 5 August 1999 (05.08.99)

(21) International Application Number: PCT/US99/01894

(22) International Filing Date: 27 January 1999 (27.01.99)

(30) Priority Data:
60/072,938 29 January 1998 (29.01.98) US

(71) Applicant *(for all designated States except US)*: THOMAS JEFFERSON UNIVERSITY [US/US]; 1020 Locust Street, Philadelphia, PA 19107 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: FEITELSON, Mark, A. [US/US]; 1717 Gwynedd View Road, North Wales, PA 19454 (US).

(74) Agent: MONACO, Daniel, A.; Seidel, Gonda, Lavorgna & Monaco, P.C., Suite 1800, Two Penn Center Plaza, Philadelphia, PA 19102 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: MOLECULAR MARKERS FOR DIAGNOSING HEPATOCELLULAR CARCINOMA

(57) Abstract

A method for diagnosing hepatocellular carcinoma is described. The method detects differential patterns of gene expression that are caused by the presence of hepatitis B virus x antigen in liver cells.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

MOLECULAR MARKERS FOR DIAGNOSING HEPATOCELLULAR CARCINOMA

Reference to Government Grant

This invention was made in the course of research sponsored by the National Institutes of Health grants CA48656 and CA66971. The U.S. Government has certain rights in the invention.

Background of the Invention

Primary hepatocellular carcinoma (HCC) is one of the most common tumors seen in certain areas of the world. In Asia and sub-Saharan Africa it has an annual incidence rate of 500 cases per 100,000 population. In the United States and Europe, HCC accounts for 1 to 2 percent of tumors seen at autopsy (Podolsky, D.K. and K.J. Isselbacher. 1994. *Harrison's Principles of Internal Medicine*, pp. 1496-1497). There are risk factors for HCC, however, that can lead to a large increase in the likelihood that tumors will develop. For example, HCC is usually associated with a cirrhotic liver, making alcoholics more likely to develop these tumors.

The increased incidence of HCC in Asian and African populations and elsewhere has been attributed to the high incidence of chronic infection with hepatitis B virus (HBV) and hepatitis C virus (HCV). These chronic infections can lead to hepatitis and cirrhosis which are the most common risk factors for HCC. The link between HBV infection and

HCC is well established. Studies in Asia have shown that the incidence of this form of cancer over time is increased 100-fold in individuals with evidence of HBV infection as compared to non-infected controls (Podolsky, D.K. and K.J. Isselbacher. 1994. *Harrison's Principles of Internal Medicine*, pp. 1496-1497). More recent work in Europe and Japan has shown that HCV is also linked to an increased risk of HCC. In fact, any agent or factor that contributes to chronic, low-grade liver cell damage would make liver cell DNA more susceptible to damage and genetic alterations which can lead to carcinogenesis. The mechanisms and steps responsible for the development of HCC, however, have not been fully elucidated.

The finding that HBV makes a genetic contribution to the development of HCC (Seeger *et al.* 1991. *J. Virol.* 65:1673-1679) suggests that one or more virus encoded proteins may play a role in hepatocarcinogenesis. Other data suggests that hepatitis B x antigen (HBxAg) contributes to the pathogenesis of HCC. HBxAg transforms a mouse hepatocyte cell line both *in vitro* and *in vivo* (Hohne, M. et al. 1990. *EMBO J.* 9:1137-1145; Seifer, M. et al. 1991. *J. Hepatol.* 13:S61-S65). HBxAg binds to and functionally inactivates the tumor suppressor p53 (Feitelson, M.A. et al. 1993. *Oncogene* 8:1109-1117; Wang, X.W. et al. 1994. *Proc. Natl. Acad. Sci. USA* 91:2230-2234; Truant, R. et al. 1995. *J. Virol.* 69:1851-1859; Takeda, S. et al. 1995. *J. Cancer Res. Clin. Oncol.* 121:593-601). HBxAg/p53 staining and complex formation has also been shown to correlate with the development of liver tumors in a X transgenic mouse model with sustained high levels of HBxAg expression (Kim, C.M. et al. 1991. *Nature* 351:317-320; Koike, K. et al. 1994. *Hepatology* 19:810-819; Ueda, H. et al. 1995. *Nature Genetics* 9:41-47).

It has previously been shown that HBxAg is a trans-activating protein (Twu, J.S. and R.H. Schloemer. 1987. *J. Virol.* 61:3448-3453; Rossner, M.T. 1992. *J. Med. Virol.* 36:101-117; Henkler, F. and R. Koshy. 1996. *J. Viral Hepatitis* 3:109-121). Even though virus DNA fragments

integrated into HCC cells often contain the X region (Matsubara, K. and T. Tokino. 1990. *Mol. Biol. Med.* 7:243-260; Unsal, H. et al. 1994. *Proc. Natl. Acad. Sci. USA* 91:822-826) and HBxAg made from these integrated sequences has transactivating-activity, it is not clear that this action is responsible for transformation (Luber, B.L. et al. 1996. *Oncogene* 12:1597-1608). A variety of studies have described differences in gene expression which distinguish tumor (HCC) form nontumor (liver) cells (Begum, N. A., M. Mori, T. Matsumata, K. Takenaka, K. Sugimachi, and G. F. Barnard. 1995. *Hepatology* 22:1447-1455; Darabi, A., S. Gross, M. Watabe, M. Malafa, and K. Watabe. 1995. *Cancer Lett.* 95:153-159; Inui, Y., S. Higashiyama, S. Kawata, S. Tamura, J.-I. Miyagawa, N. Taniguchi, and Y. Matsuzawa. 1994. *Gastroenterology* 107:1799-1804; Kim, S. O., J. G. Park, and Y. I. Lee. 1996. *Cancer Res.* 56:3831-3836; Ohmachi, Y., A. Murata, T. Yasuda, K. Kitagawa, S. Yamamoto, M. Monden, T. Mori, N. Matsuura, and K. Matsubara. 1994. *J. Hepatol.* 21:1012-1016; Su, W., J. F. Liu, S. X. Zhang, D. F. Li, and J. J. Yang. 1994. *Hepatology* 19:788-799; Uekt, T., J. Fujimoto, T. Suxukt, H. Yamamoto, and E. Okamoto. 1997. *Hepatology* 25:862-866; Yamashita, N., H. Ishibashi, K. Hayashida, J. Kudo, K. Takenaka, K. Itch, and Y. Niho. 1996. *Hepatology* 24:1437-1440; Zhou, M. X., M. Watabe, and K. Watabe. 1994. *Arch. Virol.* 134:369-378). However, no indication has been given whether any of these genes are turned on or off by HBxAg.

One of the problems associated with any type of cancer is ensuring early detection and risk factor screening so that disease can be more successfully treated. Detection of HCC may escape clinical recognition because of the presence of other active disease processes, such as hepatitis or cirrhosis. One screening tool has been alpha fetoprotein levels, where levels greater than 500 µg/L are found in 70-80% of patients with HCC (Podolsky, D.K. and K.J. Isselbacher. 1994. *Harrison's Principles of Internal Medicine*, pp. 1496-1497). The most common

diagnostic tools are imaging with ultrasound, which can only detect the presence of visible tumors, and liver biopsy. Neither of these diagnostic tools is able to screen individuals for the risk of disease before tumors develop. In biopsy, it can be difficult to distinguish large cirrhotic nodules from well-differentiated HCC or low-grade dysplastic nodules from HCC.

Clearly, there is a need for better methods of early diagnosis, as well as risk screening. Criteria for judging the usefulness of HCC screening methods were recently reviewed by Collier and Sherman, 1988. *Hepatology* 27:273-278.

Summary of the Invention

The invention is a method for detecting hepatocellular carcinoma in liver tissue of a patient. A liver tissue sample is obtained from the patient, and the level of expression of one or more marker genes in the sample is assessed. The marker genes are differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells. A reduction in the level of expression of one or more marker genes in the sample as compared to the expression level in noncancerous liver tissue is indicative of hepatocellular carcinoma in the sample.

According to an embodiment of the invention, the marker gene is selected from the group of genes expressing RNA transcripts which hybridize under conditions of high stringency to a nucleic acid probe selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10. More particularly, the marker gene is selected from the group consisting of

a gene which encodes the polypeptide of SEQ ID NO:27;

a gene which encodes the polypeptide of SEQ ID NO:28;

a gene which encodes the polypeptide of SEQ ID NO:29;

a gene which encodes the polypeptide of SEQ ID NO:29;

hu-*sui*;

human tubulin-folding cofactor E gene;

human myeloblast KIAA0132 gene; and

the human fetal heart gene, the cDNA of which is identified as GenBank accession number AA047006.

An example of high stringency hybridization conditions is hybridization at 4XSSC at 65°C, followed by washing in 0.1XSSC at 65°C for one hour. Another example of high stringency hybridization conditions is hybridization in 50% formamide, 4XSSC at 55°C.

According to one embodiment of the invention, the step of assessing the level of expression of the marker gene in the sample comprises contacting the sample with one or more probes which detect mRNA which is differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells.

According to another embodiment of the invention, the step of assessing the level of expression of the marker gene by in the sample comprises assessing the level of expression of marker protein encoded by one or more marker genes. Detection of marker protein is accomplished by contacting the sample with one or more antibodies which bind marker proteins.

According to another embodiment, the invention provides a method for diagnosing hepatocellular carcinoma comprising the steps of obtaining a liver tissue sample from a patient, and assessing the level of expression of one or more marker genes in the tissue sample, which marker genes are differentially expressed in HBxAg[+] cells as compared

with HBxAg[-] cells. The reduction of detectable expression of one or more marker genes in the sample is indicative of hepatocellular carcinoma.

In yet another embodiment, the invention provides a method for identifying patterns in gene expression in a biological sample that are altered by hepatitis B x antigen comprising the steps of

obtaining a biological sample;

contacting said sample with a probe which detects an mRNA which is differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells; and

detecting expression of a gene encoding said mRNA detected by the probe.

Alternatively, the steps for identifying alterations in gene expression patterns in the biological sample comprise

contacting said sample with an antibody which detects a protein which is differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells; and

detecting expression of a gene encoding the protein detected by the antibody.

Abbreviations and Definitions

A. Abbreviations

"ABC"	avidin-biotin-peroxidase complex
"bp"	base pair
"CAT"	chloramphenicol acetyltransferase
"ISH"	*in situ* hybridization
"TTP"	deoxythymidine triphosphate
"HBxAg"	hepatitis B x antigen
"HBsAg"	hepatitis B surface antigen
"HBV"	hepatitis B virus
"HCC"	hepatocellular carcinoma

"PCR"	polymerase chain reaction
"RT"	reverse transcriptase
"SSC"	standard saline citrate solution (0.15M saline containing 0.015M sodium citrate, pH 7)

B. Definitions

"Expression" means, with respect to a gene, the realization of genetic information encoded in the gene to produce a functional RNA or protein. The term is thus used in its broadest sense, unless indicated to the contrary, to include either transcription or translation.

"Expression level", with respect to a gene means a relative expression level as determined by comparison with the expression level of the gene in noncancerous tissue. An expression level may be "assessed" visually in a sample with the aid of a microscope, such as by considering the intensity of a stain for protein encoded by the gene of interest, or by considering the relative number of stained versus unstained cells in the sample.

"Hybridization" means the Watson-Crick base-pairing of essentially complementary nucleotide sequences (polymers of nucleic acids) to form a double-stranded molecule.

"Marker gene" means a gene which is differentially expressed in HCC tumor versus non-tumor tissue.

"Marker protein" means a protein which is encoded by a marker gene.

Detailed Description of the Invention

The present invention is a method for determining whether tissue from a biopsy represents HCC, based on detection of gene expression patterns in cells. Studies of tumor and non-tumor pairs from patients demonstrate the differential expression of certain genes in tumor versus non-tumor tissue. The genes are expressed in non-HCC tissue, but

expression is substantially reduced or undetectable in HCC tumor tissue. Thus, a reduction in expression of one or more of the marker genes in the tissue sample is diagnostic for the presence of HCC in the patient sample. At the individual cell level, one or more of the marker genes may not be expressed or may be characterized by reduced expression level, such that the expression of the gene in the tissue sample as a whole is reduced. The identification of such molecular markers provides a method for diagnosing HCC without relying on tissue morphology alone. This is the first time that molecular markers associated with chronic HBV infection have been shown to be useful in the diagnosis of HCC.

The diagnostic marker genes were identified by manipulation of HepG2X cells. HepG2 is a differentiated cell line derived from a human hepatoblastoma. The cell line HepG2X was generated by infection of HepG2 cells by replication defective recombinant retroviruses encoding the full length HBxAg polypeptide. HepG2CAT cells were generated in the same manner by substituting the bacterial CAT gene for the HBV X gene in the transfection vector. The HepG2X cells express the HBV X antigen (HBxAg[+]), while HepG2CAT cells do not (HBxAg[-]).

Genes whose expression were either turned on or off in the presence of the hepatitis B x antigen (HBxAg) in HepG2 cells are identified by PCR select cDNA subtraction. Briefly, the method consists of isolating whole cell RNA from HBxAg [+] and [-] HepG2 cells. Methods and kits for performing PCR select cDNA subtraction are well-known and commercially available, e.g., from Clontech, Palo Alto, CA. The RNA from HepG2X cells is subtracted from those in HepG2 cells, providing RNAs expressed in HepG2X cells but not in HepG2 cells. The RNAs were then reverse transcribed into DNA and then PCR amplified using random primers. In order to obtain RNAs expressed in HepG2, but not HepG2X cells, the opposite subtraction is carried out. These RT/PCR fragments were then cloned and either partially or fully sequenced.

Accordingly, equivalent amounts of poly(A)$^+$ RNA were isolated from confluent cultures of HepG2X and HepG2CAT cells and subjected to PCR select cDNA subtraction. DNA strands were individually sequenced from every clone, and the results for each compared to entries in GenBank and other related databases (Table 1, below). The PCR select cDNA subtraction generated gene fragments from up to eight different cellular genes that were detected in HepG2X cells but not in HepG2CAT cells (L4, L7, L8, L11, L12, L15, L16 and L19). Five of these (L7, L8, L12, L16 and L19) had at least 89% homology with fragments of known products from GenBank (Table 1). Interestingly, three of the five sequences (L7, L12, and L19) had homology with factors upregulated in fetal tissues, suggesting that they may have some growth regulatory functions. In addition, two fragments (C1 and C2) were apparently present in HepG2CAT cells but absent in HepG2X cells. Hence, up to ten genes were differentially expressed in HepG2X compared to HepG2CAT cells. In the case of the transcripts hybridizing to the L fragments, the clones represent fragments of genes whose expression is activated in HBxAg[+] cells compared to HBxAg[-] cells. In the case of transcripts hybridizing to the C fragments, the clones represent genes whose expression is suppressed in HBxAg[+] cells compared to HBxAg[-] cells. The fragment size given in Table 1 is considered approximate, as size was estimated visually from gels.

Table 1: Differentially expressed genes in HBxAg[+] and [-] Hep G2 cells

clone	insert size (~bp)	GenBank Search Match (and Accession #)	%homology in overlap
HBx[+][a] Clones			
L7	690	human fetal liver cDNA clone (H49417)	95.7% in 440 bp
L8[b]	220	human tubulin-folding cofactor E cDNA (U61232)	100% in 45 bp
L12	320	human 40S ribosomal protein S15A (P48149)	-
L16[b]	180	human myeloblast KIAA0132 gene	99% in 65 bp
L19	250	human fetal heart cDNA (AA026758)	99% in 152 bp
L4[b]	1700	none	-
L11	580	none	-
L15	1580	none	-
HBx[-][a] Clones			
C2[b]	620	human *sui1* (L26247)	-
C1	670	none	-

[a]The clones represent fragments of genes whose expression is activated (L4,L7,L8,L11,L12,L15,L16,L19) or suppressed (C1,C2) in HBxAg[+] compared to HBxAg[-] cells.

[b]Probes whose sequences share considerable homology with sequences independently found in tumor compared to nontumor cells.

The cDNA fragments obtained from subtraction hybridization (Table 1) were used as probes for ISH of HepG2X and HepG2CAT cells, to verify that the probes obtained from PCR select cDNA subtraction actually represented differentially expressed genes in HepG2 compared to HepG2X cells. In all cases, the L probes hybridized to HepG2X cells. Little or no

signal was observed in HepG2CAT cells. In contrast, the C probes demonstrated strong hybridization in HepG2CAT cells, but little or no signal in HepG2X cells. Thus, in all cases, the probes obtained from PCR select cDNA subtraction actually reflected differences in gene expression between HepG2X and HepG2CAT cells.

The cDNA fragments were either partially (L4, L7, L8, L15, L16 and C2) or completely (L11, L12, L19 and C1) sequenced. The sequences are as follows:

Table 2: Nucleotide Sequences of Fragments of Differentially Expressed cDNA

cDNA Fragment	Fragment Nucleotide Sequence
L7	SEQ ID NO:1
L8	SEQ ID NO:2
L12	SEQ ID NO:3
L16	SEQ ID NO:4
L19	SEQ ID NO:5
L4	SEQ ID NO:6
L11	SEQ ID NO:7
L15	SEQ ID NO:8
C2	SEQ ID NO:9
C1	SEQ ID NO:10

In order to further study the structure and function of the protein encoded by the C2 mRNA, the full length cDNA containing the C2 sequence was obtained (from HepG2CAT cells) by 5' and 3' rapid amplification of cDNA ends (RACE) PCR using the Marathon™ cDNA Amplification Kit (Clontech, Palo Alto, CA). Briefly, one 3' and one 5' gene specific primers were synthesized. PCR was performed using these primers together with an adaptor primer to obtain the 3' or 5' cDNA specific products in separate amplification reactions. The products were cloned into pT7Blue T (Novagen, Inc., Madison, WI) and sequenced. The

appropriate 3' and 5' gene specific fragments were then digested with suitable restriction enzymes and cloned into pcDNA3 (Invitrogen, San Diego, CA) at the chosen site(s), and the integrity of the full length clone verified by DNA sequencing. This resulted in a full length clone exactly 1.35 kb in length, which encoded a small protein of 113 amino acids near its 5' end that has 100% homology with the human translation initiation factor, hu-*sui1*. The C2 probe spans bases 903-1350 of full length hu-*sui1* cDNA

Other than its regulatory role in translation initiation, the human *sui1* protein does not appear to have any recognizable motifs which would suggest additional functions. These results indicate that the introduction of HBxAg results in the altered expression of a protein whose function is associated with the regulation of translation. Further, HBxAg may contribute to hepato-carcinogenesis, in part, by altering gene expression at the level of translation initiation.

Additional full length cDNAs from differentially expressed genes containing fragments L4, L7, L11 and L12 were obtained in a similar manner to fragment C2. The cDNA containing fragment L12 encoded a protein of 130 amino acids having a 100% homology with the human 40S ribosomal protein S15A (Accession nos. P39027, P39031). Sequences of the full length cDNAs and corresponding gene names are set forth in Table 3:

Table 3: Nucleotide Sequences of Full-length Differentially Expressed cDNAs

cDNA Fragment	Full Length cDNA Seq. (SEQ ID NO:)	Gene Name	GenBank Accession Number	Translated Protein Seq. (SEQ ID NO:)
L4	11	unknown	-	27
L7	12	human fetal liver cDNA	H49417	28
L11	13	unknown	-	29
L12	14	human 40S ribosomal protein S15A	P48149	30
C2	15	hu-*sui*1	L26247	31

Experiments were performed to detect hu-*sui1* transcripts in tumor and nontumor tissues from HBV infected patients. A panel of tumor/nontumor tissue pairs from a group of patients were analyzed by ISH using the C2 probe. Among this group, 14 patients were from South Africa, while the remaining 23 were from mainland China. The results (Table 5, Example 3, below) show that hu-*sui1* mRNA is easily detectable in nontumor tissue from both groups, but that it is rarely present in tumor tissues from the same patients Thus, hu-*sui1* is differentially expressed in tumor vs. non-tumor tissue.

ISH was performed with the full set of ten individual probes (L4, L7, L8, L11, L12, L15, L16, L19, C1 and C2) on tumor/nontumor paired samples from five HBV carriers with HCC, and on normal uninfected liver from two individuals. The probes detected transcripts that were

preferentially expressed in nontumor, compared to tumor tissues, in most cases (Table 6, Example 4, below). These results were not due to differences in the ability of the tumor tissue to uptake probe, since tumor cells from three of the five HCC patients hybridized strongly to an alpha fetoprotein probe.

For diagnosis of HCC, a sample of liver tissue is removed from an individual by conventional biopsy techniques which are well-known to those skilled in the art. Typically, the test subject will be an HBV-infected individual. The sample is generally collected by needle biopsy. Procedures for liver needle biopsy are well-known in medicine. A mass may be apparent from either tactile examination of the patient, or upon imaging such as by ultrasound. The needle biopsy should be taken at or near the site of the mass. Ultrasound guided percutaneous fine-needle biopsy procedures are known. *See, e.g.*, Polakow, *Hepatogastroenterology* 45:1829-30 (1998). The biopsy sample may also be taken in connection with a surgical procedure in which the liver becomes accessible.

The expression level of the marker gene may serve as a convenient molecular marker to replace or augment conventional liver tissue examination, which largely relies on subjective criteria. This form of "molecular-based" diagnosis can be performed more consistently than conventional pathological examination which is based upon subjective evaluations by expert pathologists.

Detecting the expression of the marker gene in the tissue sample comprises detecting RNA transcripts, particularly mRNA transcripts in the sample tissue, or detecting the corresponding marker gene product (protein) in the sample tissue. Preferably, the presence of the marker protein in the sample tissue is detected by an immunoassay whereby an antibody which binds the marker protein is contacted with the sampled tissue.

Typically, a portion of noncancerous liver tissue will be removed with the purported tumor tissue during biopsy. The noncancerous (i.e., non-tumor) cells will express the marker gene, and will provide a positive signal for the absence of HCC. Hence, the noncancerous cells in the biopsied sample will serve as a convenient positive control.

It may be desirable in some cases to compare the assay results against a control sample comprising liver cells from non-tumor liver tissue from the test subject, or non-tumor liver tissue from another (HBV-infected) individual. The non-tumor sample should test positive for the expression of the marker gene.

Tissue samples may be considered as HCC-positive when the level of expression of one or more marker genes is reduced in the tissue sample as a whole, compared to the expression level in noncancerous liver tissue. The overall reduction of marker gene expression in the liver sample may arise from a reduced but still detectable expression level in at least a portion of the cells of the sample, a complete loss of marker gene expression in some cells, or a combination of both. The former may be apparent as a general lessening of stain intensity when the sample is treated with a stain for cells which express the marker gene. The latter may be observed as a complete absence of stain in the affected cells. These observations can be made visually with the aid of a microscope.

Methods of detecting mRNA transcripts of a particular gene in cells of a tissue of interest are well-known to those skilled in the art. According to one such method, total cellular RNA is purified from the effected cells by homogenization in the presence of nucleic acid extraction buffer, followed by centrifugation. Nucleic acids are precipitated, and DNA is removed by treatment with DNase and precipitation. The RNA molecules are then separated by gel electrophoresis on agarose gels according to standard techniques, and transferred to nitrocellulose filters by, *e.g.*, the so-called "Northern" blotting technique. The RNA is immobilized on the filters by

heating. Detection, and quantification if desired, of specific RNA is accomplished using appropriately labeled DNA or RNA probes complementary to the RNA in question. See *Molecular Cloning: A Laboratory Manual*, J. Sambrook *et al.*, eds., 2nd edition, Cold Spring Harbor Laboratory Press, 1989, Chapter 7, the disclosure of which is incorporated by reference.

More preferably, the mRNA assay is carried out according to ISH. Also known as "cytological hybridization", the *in situ* technique involves depositing whole cells or tissues onto a microscope cover slip and probing the nucleic acid content of the cell with a solution containing radioactive or otherwise labeled cDNA or cRNA probes. The practice of ISH is described in more detail in U.S. Patent 5,427,916, the entire disclosure of which is incorporated herein by reference.

The nucleic acid probes for the above RNA hybridization methods can be designed based upon the full length marker gene sequences described or referenced herein. Where the marker gene has yet to be identified with a known, full-length sequenced DNA, the corresponding cDNA fragment listed in Table 2 may be used as the probe.

Methods for preparation of labeled DNA and RNA probes, and the conditions for hybridization thereof to target nucleotide sequences, are described in *Molecular Cloning, supra*, Chapters 10 and 11, incorporated herein by reference. The nucleic acid probe may be labeled with, *e.g.*, a radionuclide such as ^{32}P, ^{14}C, or ^{35}S; a heavy metal; or a ligand capable of functioning as a specific binding pair member for a labeled ligand, such as a labeled antibody, a fluorescent molecule, a chemolescent molecule, an enzyme or the like.

Probes may be labeled to high specific activity by either the nick translation method or Rigby *et al.*, *J. Mol. Biol.* 113: 237-251 (1977) or by the random priming method, Fienberg *et al.*, *Anal. Biochem.* 132: 6-13 (1983). The latter is the method of choice for synthesizing ^{32}P-labeled

probes of high specific activity from single-stranded DNA or from RNA templates. Both methods are well-known to those skilled in the art and will not be repeated herein. By replacing preexisting nucleotides with highly radioactive nucleotides, it is possible to prepare ^{32}P-labeled DNA probes with a specific activity well in excess of 10^8 cpm/microgram according to the nick translation method. Autoradiographic detection of hybridization may then be performed by exposing filters on photographic film.

Where radionuclide labeling is not practical, the random-primer method may be used to incorporate the TTP analogue 5-(N-(N-biotinyl-epsilon-aminocaproyl)-3-aminoallyl)deoxyuridine triphosphate into the probe molecule. The thus biotinylated probe oligonucleotide can be detected by reaction with biotin binding proteins such as avidin, streptavidin, or anti-biotin antibodies coupled with fluorescent dyes or enzymes producing color reactions.

In situ hybridization is most conveniently carried out using a commercially available kit for labeling nucleic acid probes with, e.g. digoxigenenin/biotin as a label. One such kit is available from Oncor, Gaithersburg, MD.

According to another embodiment of the invention, marker gene expression in cells of the patient tissue is determined by detecting the corresponding marker protein. A variety of methods for detecting and quantifying expression of proteins of interest exist, including Western blotting and immunohistochemical staining. The latter is preferred. Western blots are run by spreading a protein sample on a gel, using an SDS gel, blotting the gel with a cellulose nitrate filter, and probing the filters with labeled antibodies. With immunohistochemical staining techniques, a tissue sample is prepared, typically by dehydration and fixation, followed by reaction with labeled antibodies specific for the desired gene product. The antibodies may be coupled to a visually detectable label, such as enzymatic labels, flourescent labels, luminescent labels, and the like.

According to one embodiment of the invention, tissue samples are obtained from patients and the samples are embedded and serially sectioned at 3-5 μ per section. The sections are fixed, mounted and dried according to conventional tissue mounting techniques. The fixing agent may advantageously comprise formalin. The embedding agent for mounting the specimen may comprise, *e.g.*, paraffin. The samples may be stored in this condition.

Following deparaffinization and rehydration, the samples are contacted with an immunoreagent comprising an antibody specific for a marker protein of interest. The antibody may comprise a polyclonal or monoclonal antibody. The antibody may comprise an intact antibody, or fragments thereof capable of specifically binding marker protein. Such fragments include, but are not limited to, Fab and $F(ab')_2$ fragments. As used herein, the term "antibody" includes both polyclonal and monoclonal antibodies. The term "antibody" means not only intact antibody molecules, but also includes fragments thereof which retain antigen binding ability.

Appropriate polyclonal antisera may be prepared by immunizing appropriate host animals with marker protein and collecting and purifying the antisera according to conventional techniques known to those skilled in the art. Monoclonal antibody may be prepared by following the classical technique of Kohler and Milstein, *Nature 254*:493-497 (1975), as further elaborated in later works such as *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analysis*, R. H. Kennet *et al.*, eds., Plenum Press, New York and London (1980).

Substantially pure marker protein for use as an immunogen for raising polyclonal or monoclonal antibodies may be conveniently prepared by recombinant DNA methods.

As an alternative to immunization with the complete marker protein, antibody against marker proteins can be raised by immunizing appropriate hosts with immunogenic fragments of the whole protein, particularly

peptides corresponding to probable antigenic determinants. Hydrophilic regions, which face the environment surrounding the protein, are most likely to contain antigenic sites. Such regions can be identified using standard computer programs.

The antibody either directly or indirectly bears a detectable label. The detectable label may be directly attached to the primary anti-marker protein antibody. More conveniently, the detectable label is attached to a secondary antibody, *e.g.*, goat anti-rabbit IgG, which binds the primary antibody. The label may advantageously comprise, for example, a radionuclide in the case of a radioimmunoassay; a fluorescent moiety in the case of an immunofluorescent assay; a chemiluminescent moiety in the case of a chemiluminescent assay; or an enzyme which cleaves a chromogenic substrate, in the case of an enzyme-linked immunosorbent assay.

Most preferably, the detectable label comprises an avidin-biotin-peroxidase complex (ABC) which has surplus biotin-binding capacity. *See* Hsu *et al.*, *J. Histochem. Cytochem.* 29:577-580, 1981. The secondary antibody is biotinylated. Kits for staining proteins by the ABC method are commercially available (e.g., Vector Laboratories, Burlingame, CA).

To determine the presence of marker protein antigen in the tissue section under analysis, the section is treated with primary antiserum against the antigen, washed, and then treated with the secondary antiserum. The subsequent addition of ABC localizes peroxidase at the site of the specific antigen, since the ABC adheres non-specifically to biotin. Peroxidase (and hence antigen) is detected by incubating the section with *e.g.* H_2O_2 and diaminobenzidine (which results in the antigenic site being stained brown) or H_2O_2 and 4-chloro-1-naphthol (resulting in a blue stain).

The ABC method can be used for paraffin-embedded sections, frozen sections, and smears. Endogenous (tissue or cell) peroxidase may be quenched *e.g.* with H_2O_2 in methanol.

The level of marker protein expression in samples may be compared on a relative basis to the expression in non-tumor liver tissue samples by comparing the stain intensities, or comparing the number of stained cells. The lower the stain intensity in the test sample with respect to nontumor controls, or the lower the stained cell count in a tissue section having approximately the same number of cells, the lower the expression of the marker gene in the sample, which indicates the presence of HCC in the sample. If a control is utilized, it advantageously comprises non-tumor liver tissue from another HBV-positive individual or non-tumor liver tissue from the patient.

As a further control of the protein expression, one may preincubate the immune serum raised against the marker protein antigen with the relevant peptide antigen. This should dissipate the signal from the immune serum when contacted with healthy or non-tumor cells, confirming that the immune serum reagent is indeed specific for the target antigen.

The diagnostic procedure described herein may take the form of detecting the expression of just one of the marker genes. Alternatively, a mixture of probes (nucleic acid or antibody) targeting different marker genes may be utilized. This may be achieved by pooling two or more nucleic acid probes in the case of a nucleic acid hybridization assay, or by pooling two or more antisera in the case of a protein assay. By testing for the expression of multiple marker genes in this manner, it is expected that the sensitivity of the assay will be increased.

The following nonlimiting examples are provided to better illustrate the claimed invention.

Example 1

Preparation of HBxAg[+] and HBxAg[-] Cell Lines

A. Cell Lines and Culture Conditions

HepG2 cells, a differentiated cell line derived from a human hepatoblastoma (Aden,D.P. et al. 1979 *Nature* 282:615-617; Knowles, B.B. et al. 1980. *Science* 209:497-499), were cultured on type-1 rat tail collagen (Becton Dickinson, Franklin Lakes, NJ) coated tissue culture dishes or plates. Cells were grown in Earle's MEM supplemented with 10% heat inactivated fetal calf serum (FCS), 100 µM MEM non-essential amino acids, 1 mM sodium pyruvate, as well as standard concentrations of penicillin plus streptomycin. The retrovirus packaging cell line PA317 (Danos, O. 1991. *Methods in Molecular Biology, Practical Molecular Virology: Viral Vectors for Gene Expression* 8:17-27) was also grown on plastic dishes in the same medium.

B. Plasmid Construction

The retroviral vector plasmid, pSLXCMVneo, was used to clone the HBV X gene (Valenzeula, P. et al. 1980. *Animal Virus Genetics*, Academic Press: New York, pp. 57-70) or the bacterial chloramphenicol acetyltransferase (CAT) gene sequences for these studies, as described (Duan, L.X. et al. 1995. *Human Gene Ther.* 6:561-573). Briefly, pSLXCMV-CAT was constructed by inserting a 726 bp HindIII-BamHI fragment containing the CAT gene into the HpaI-BglII site of the pSLXCMV polylinker. PSLXCMV-FLAG-HBx was constructed by inserting a 920 bp MluI-BglII fragment of FLAG-HBx DNA into the MluI-BglII site of the pSLX-CMV polylinker. Recombinants were used to transform HB101. Minipreps were prepared and the DNA used for sequence analysis.

C. Preparation of Recombinant Retroviruses and Infection of HepG2 Cells

Approximately 1 x 10^6 PA317 cells/100 mm dish were transfected using standard calcium phosphate precipitation using 15 µg of pSLXCMV-FLAG-HBx or 15 µg of pSLXCMV-CAT. At 24, 48, and 72 hours after transfection, the medium was removed and processed through a 0.45 µm filter to remove PA317 cells, and then used immediately for infection of HepG2 cells. Five ml of recombinant retrovirus-enriched supernatant (5 x 10^5 CFU/ml, as assayed on NIH-3T3 cells) was used to infect 1 x 10^6 target HepG2 cells/100 mm dish in the presence of polybrene (8 µg/ml) for 24 hours. Fresh virus supernatant was added after 24 and again after 48 hours so that the cells were exposed to virus for a total of 72 hours. All of these infections were carried out in log phase cultures. Cells were then passaged at 1:2 and selected by incubation in G418 (800 µg/ml; GIBCO/BRL, Grand Island, NY) for 14 days in order to maximize the fraction of cells producing HBxAg or CAT. G418 colonies were then expanded in normal growth medium and used for analysis. The fourteen day selection in G418 had the effect of eliminating most of the uninfected cells.

D. Detection of CAT Activity and HBxAg Polypeptide in Transfectants

The transfectants (HepG2-CAT and HepG2X) were evaluated as follows.

CAT assays were performed as described by Wang et al. (1994. *Proc. Natl. Acad. Sci. USA* 91:2230-2234). Briefly, 1 x 10^7 HepG2-CAT cells in a 100 mm dish were lysed by addition of 0.9 ml of 1x report lysis buffer (Promega) for 15 minutes and harvested by scraping. Cells were pelleted and 180 µl of cell lysate was used for a standard CAT assay. After incubation with ^{14}C-chloramphenicol, acetylated forms were separated by thin-layer chromatography. Alternatively, lysates prepared from 5 x 10^6

HepG2X cells were assayed for the 17 kDa HBxAg by western blotting using a mixture well characterized rabbit anti-x peptide antibodies (Feitelson, M.A. and M.M. Clayton. 1990. *Virology* 177:367-371; Feitelson, M.A. et al. 1990. *Gastroenterology* 98:1071-1078). Horseradish peroxidase conjugated goat anti-rabbit Ig (Accurate, Westbury, NY) and ECL substrate (Amersham, Arlington Heights, IL) were used for detection.

CAT activity was present in HepG2CAT, but not in HepG2X cells. HBxAg was present in lysates from HepG2X, but not from HepG2CAT cells. Together, these findings show that both of the recombinant retroviruses are expressing the expected products in HepG2 cells.

Example 2

Identification of Differentially Expressed Genes Distinguishing HepG2X from HepG2CAT

The differences in gene expression which distinguish HepG2X from HepG2CAT cells were determined by using a commercially available subtraction hybridization approach (the PCR-select cDNA subtraction kit from Clontech, Palo Alto, CA). Briefly, whole cell RNA was extracted separately from 1×10^7 HepG2X and an equal number of HepG2CAT cells, and the quality of the extraction was determined by assaying for 18S and 28S rRNAs by agarose gel electrophoresis and ethidium bromide staining. PCR-select cDNA subtraction is reverse transcriptase (RT)/PCR based, and enriches for poly A^+ RNA (isolated using the Qiagen RNeasy total RNA kit; QIAGEN, Inc., Chatsworth, CA) from tissue culture cells or tissues. The procedure involved ligating adaptors to some of the PCR products and conducting two rounds of subtractive hybridization against the PCR products from the cells in which the comparison were being made. The resulting products were then PCR amplified using primers which matched

the sequence of the adaptors (in the CLONTECH Advantage cDNA PCR kit). The unique fragments were then eluted from the gels (using the QIAGEN gel extraction kit) and cloned into pT7Blue (Novagen, Madison, WI). Positive clones were selected by blue-white phenotype. Recombinant DNAs were isolated from minipreps of individual clones, digested by *Rsa I* to check insert size, and then both strands individually analyzed by sequence analysis. The sequences obtained were then compared to those in GenBank using the FASTA command in the GCG software package for homology to known genes. The results are set forth in Table 1, above.

Example 3

Detection of *Hu-sui1* Differential Expression in Patient Tumor and Non-tumor Tissue by *In situ* Hybridization

These experiments were performed to detect hu-*sui1* transcripts in a tumor and nontumor tissues from HBV infected patients. Accordingly, a panel of tumor/nontumor tissue pairs from a group of HCC patients was analyzed by ISH using the C2 probe.

The HCC and surrounding nontumor liver tissues were obtained from two different sets of HCC patients. The characteristic of the patients are set forth in Table 4:

Table 4: Characteristics of HCC patients in study

Patient group	*African*	*Chinese*
Number tested	14	23
Race	13 black, 1 Caucasian	23 Chinese
Gender	14 male, 0 female	17 male, 6 female
Age range:	14-72 years	31-68 years
mean:	39 years	48 years
No. $HBsAg^+$ of total tested	8 of 14 tested (57%)	9 of 23 tested (39%)
No $HBeAg^+$ of total tested	2 of 13 tested (15%)	14 of 23 tested (61%)
No. anti-HBc^+ of total tested	13 of 14 tested (93%)	19 of 23 tested (83%)
No anti-HBc^+ of total tested	2 of 13 tested (15%)	10 of 23 tested (43%)
No. anti-HBs^+ of HBs tested	5 of 6 HbsAg [-] cases	6 of 14 HbsAg [-] cases

In Table 4, HBsAg and HBcAg are, respectively, hepatitis B surface antigen and core antigen. HBeAg is hepatitis B e antigen, a proteolytic fragment of HBcAg which is secreted as a free polypeptide into the blood of patients who replicate virus in the liver. HBeAg has thus been described as a surrogate marker for virus replication

Twenty-three paired tumor/nontumor samples came from as many HBsAg positive Chinese carriers who had undergone surgery for the removal of their tumors. Most patients lived in and around Xi'an, China and were treated at the Fourth Military Medical University. Fourteen additional paired tumor/nontumor samples from as many patients were obtained from South African patients. Half of these were HBV carriers (serum HBsAg positive) while the remaining patients, except for one, had evidence of past HBV infection (detectable anti-HBs and/or anti-HBc). Formalin fixed, paraffin embedded tissues, fresh frozen blocks, and -80°C snap frozen paired liver and tumor samples from individual patients were collected from most patients, used for diagnostic purposes, and were then made available for these studies. Analogous pieces of uninfected human liver from two individuals were available to serve as controls.

Gene fragment C2 obtained from PCR select cDNA subtraction was used as a probe in *in situ* hybridization using the Oncor ISH and digoxigenenin/biotin detection kits according to the instructions provided by

the manufacturer (Oncor, Gaithersburg, MD). The results, shown in Table 5, demonstrate that hu-*sui1* mRNA is easily detectable in nontumor tissue from both groups, but that it is rarely present in tumor tissues from the same patients. For example, 13 of 14 South African patients (93%) and 22 of 23 Chinese patients (96%) had detectable hu-*sui1* mRNA by ISH in nontumor cells. In contrast, only 1 South African (7%) and 5 Chinese (22%) had detectable hu-*sui1* mRNA by ISH in tumor tissue. Among the Chinese patients with detectable hu-*sui1* in HCC, 3 of the 5 had only trace amounts of signal in less than 10% of the tumor cells. Nontumor tissue signals were often more intense and more widespread. These patterns were observed in both HBsAg positive and negative patients with HCC in both ethnic groups. These results demonstrate that hu-*sui1* is differentially expressed in tumor compared to nontumor tissue.

Table 5: Summary of *in situ* hybridization for C2 probe in tumor /nontumor pairs for HCC patients from south Africa and China[a]

South Africa patients

case no:	1	2	3	4	5	6	7	
tumor	0	0	0	0	2	0	0	
nontumor	3	3	3	3	4	3	3	
case no:	8	9	10	11	12	13	14	
tumor	0	0	0	0	0	0	0	
nontumor	3	2	4	3	2	0	3	

Chinese patients

case no:	1	2	3	4	5	6	7	8
tumor	2	0	1	0	0	0	0	1
nontumor	3	2	4	3	3	4	3	3
case no:	9	10	11	12	13	14	15	16
tumor	0	0	0	0	1	0	0	0
nontumor	4	0	4	4	3	3	3	4
case no:	17	18	19	20	21	22	23	
tumor	0	0	2	0	0	0	0	
nontumor	3	3	3	3	4	4	4	

[a]*In situ* hybridization (ISH) staining is estimated as follows: **0**: no signal; **1**: ISH signal in <10% of cells; **2**: ISH signal in 10-25% of cells; **3**: ISH signal in 25-50% of cells; **4**: ISH signal in >50% of cells.

Example 4

Detection of Differential Expression in Patient Tumor and Non-tumor Tissue by *In situ* Hybridization

The HCC and surrounding nontumor liver tissues used for analysis were obtained from five HBsAg positive Chinese carriers who had undergone surgery for the removal of their tumors. These patients were treated at the Fourth Military Medical University, Xi'an, China. Formalin fixed, paraffin embedded tissues, fresh frozen blocks, and -80°C snap

frozen paired liver and tumor samples from individual patients were collected, used for diagnostic purposes, and were then made available for these studies. In many cases, tumor was dissected from nontumor just prior to snap freezing. Analogous pieces of uninfected human liver from two individuals were available to serve as controls.

In situ hybridization was carried out using probes L4, L7, L8, L11, L12, L15, L16, L19, C1 and C2, and the Oncor ISH and digoxigenenin/biotin detection kits according to the instructions provided by the manufacturer (Oncor, Gaithersburg, MD). The results are shown in Table 6. The probes detected transcripts that were preferentially expressed in nontumor, compared to tumor tissues, in most cases. These results were not due to differences in the ability of the tumor tissue to uptake probe, since tumor cells from three of the five HCC patients hybridized strongly to a 320 bp alphafetoprotein probe (data not shown). Hence, the probes that distinguish HepG2X from HepG2CAT cells also distinguish tumor from nontumor in carriers with HCC.

Table 6: *In situ* hybridization results from PCR select cDNA amplification

clone	hybridization[a] in		tumor/nontumor (T/NT) in patient:					uninfected
	HepG2 -X	HepG2 -CAT	1 T NT	2 T NT	3 T NT	4 T NT	5 T NT	liver
L7	++	±	+/+++	+/+++	-/+++	-/+	±/+	++
L8	++	-	-/++	-/++	-/+	+/++	-/++	+
L12	++	±	-/+	+/++	-/++	-/++	-/++	±
L16	+	-	±/++	-/+	-/-	±/+	+/++	+
L19	++	+	-/++	-/++	+/++	-/+	-/++	±
L4	++	±	-/+++	++/-	+++	±/+	-/+	±
L11	+	-	-/+++	-/+++	±/++	-/+	-/++	+
L15	++	+	+/++	-/++	+/++	-/++	-/+++	+
C2	-	++	-/++	-/+	-/+++	-/+	-/+	+++
C1	±	++	-/++	-/++	-/+++	+/+++	+/++	+++

[a]*In situ* hybridization (ISH) signals were as follows: - : no signal; ± : 1-10% of the cells were positive for the corresponding probe; + : 11-25% of the cells were positive; ++ : 26-50% of the cells were positive; +++ : >50% of the cells positive. In the great majority of cases, HbxAg was observed in nontumor liver. Tumor cells were either faintly positive for HbxAg or completely negative.

Example 5

Detection of Differential Expression in Patient Tumor and Non-tumor Tissue by Immunostaining

A. Peptides

Synthetic peptides that represent probable antigenic determinants on each of the differentially expressed proteins were prepared by solid

phase peptide synthesis and analyzed by HPLC and amino acid composition prior to use. The peptides are identified in Table 7, below. The The peptides were coupled by virtue of their free cysteine sulfhydryl (either in the peptide sequence or added to the carboxy or amino terminus where the native sequence did not contain a cysteine) to keyhole limpet hemocyanin (KLH; Sigma) using the coupling agent m-maleimidobenzyol-N-hydroxysuccinimide ester (MBS; Pierce) as described by Liu *et al.*, *Biochemistry* 18:690-697 (1979).

Table 7: Peptide fragments of differentially expressed proteins

Gene	Peptide	Sequence	Peptide Position in Protein
C2	C2.1	DDYDKKKLVKAFKKKFAC (SEQ ID NO:16)	52-69
	C2.2	EHPEYGEVIQLQGDQRKNIC (SEQ ID NO:17)	75-94
L4	L4A	CQKAKDRMERITRKIKDSDAYRRDE (SEQ ID NO:18)	460-484
	L4B	PRPRDKRQLLDPPGDLSRC (SEQ ID NO:19)	821-838
L7	L7A	CGVWNQTEPEPAATS (SEQ ID NO:20)	12-25
	L7B	HHHGRGYLRMSPLFKC (SEQ ID NO:21)	56-70
L11	1L11	PCPELACPREEWRLGP (SEQ ID NO:22)	2-17
	3L11	DPSRSPHSTSSFPRGSSATSCDSR (SEQ ID NO:23)	316-339
	4L11	HPPDGSFSTFHDGPQPLEDPC (SEQ ID NO:24)	359-378
L12	L12.1	KSINNAEKRGKRC (SEQ ID NO:25)	12-23
	L12.2	DHEERRRKHTGGKC (SEQ ID NO:26)	112-124

B. Antibody Production

For antibody production, 5- to 10-week old female New Zealand White rabbits (2 animals/peptide; Hazelton) were bled and then injected with peptide conjugate as described (Bittle *et al.*, *Nature* 298:30-33, 1982). Dilutions of immune sera were assayed in parallel with preimmune sera in solid-phase assays in wells (Immunolon 2 Removawell Strips, Dynatech Labs) coated with the appropriate (unconjugated) synthetic peptide. See Feitelson *et al.*, *Gastroenterology* 98: 1071-1078, 1990 or Feitelson *et al.*, *J Med Virol* 24:121-136, 1988 for additional details of the solid phase assay design.

C. Immunohistochemical Staining

Antisera generated from two or three peptide antigens of the same protein (Table 7) are pooled, and used in immunohistochemical staining assays as follows. Paired tumor/nontumor tissue samples from HBV-associated liver cancer HCC patients comprise the test samples.

Tissues are fixed in 10% formalin, embedded in paraffin and serially sectioned at 5m per section. Sections are then stained for individual differentially expressed proteins by the avidin-biotin complex (ABC) method (Hsu *et al.*, *J. Histochem Cytochem* 29: 577-580) using a kit purchased from Vector Laboratories (Burlingame, CA). Staining is detected by addition of diaminobendizine (DAB) substrate, and the sections then counterstained with Mayer's hematoxylin. The degree of positive reaction is scored from 0 to +++. The grade 0 indicates no demonstrable antigen, + mild, ++ moderate and +++ dark staining.

D. Results - C2.1/C2.2 Antibody Mixture

Staining with antisera comprising antibodies against the C2.1 and C2.2 antigens indicated the presence of the Sui1 protein by brown color in the cytoplasm of nontumor cells surrounding the HCC tumor tissue. Very little staining appeared in the tumor cells of the HCC tissue.

D. Controls

The specificity of staining for differentially expressed protein was demonstrated by the following controls. (a) Several uninfected human liver samples, and several tissue types from other organs (spleen, lymph node, muscle, nerve, and gall bladder), were tested with immune sera. (b) Preimmune and normal rabbit sera were tested with positive liver sections. (c) The synthetic peptide(s) used to raise immune sera was tested for blocking of staining when preincubated with corresponding antisera prior to staining. (d) Liver powder made from uninfected human liver tissues was used to absorb the primary antibodies prior to staining. Peptide antisera were tested by western blotting with *E. coli* lysate from bacteria expressing the corresponding L or C polypeptide compared to a similar lysate from untransfected host cells. The results of each of these control procedures supported the specificity of staining.

All references cited with respect to synthetic, preparative and analytical procedures are incorporated herein by reference. All sequence records identified by GenBank accession numbers are incorporated herein by reference.

The entire disclosure of U.S. provisional patent application Serial No. 60/072,938 filed January 29, 1998, is incorporated herein by reference.

The present invention may be embodied in other specific forms without departing from the spirit or essential attributes thereof and, accordingly, reference should be made to the appended claims, rather than to the foregoing specification, as indication the scope of the invention.

What is claimed is:

1. A method for detecting hepatocellular carcinoma in liver tissue of a patient comprising:

assessing the level of expression of one or more marker genes in a liver tissue sample from the patient, which marker genes are differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells, a reduction in the level of expression of said one or more marker genes in the sample as compared to the expression level in noncancerous liver tissue being indicative of hepatocellular carcinoma in the sample.

2. The method according to claim 1 wherein the one or more marker genes is selected from the group of genes expressing an RNA transcript which hybridizes under conditions of high stringency to a nucleic acid probe selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10.

3. The method according to claim 1 wherein the marker gene is selected from the group consisting of

a gene which encodes the polypeptide of SEQ ID NO:27;
a gene which encodes the polypeptide of SEQ ID NO:28;
a gene which encodes the polypeptide of SEQ ID NO:29;
a gene which encodes the polypeptide of SEQ ID NO:29;
hu-*sui*;
human tubulin-folding cofactor E gene;
human myeloblast KIAA0132 gene; and
the human fetal heart gene, the cDNA of which is identified as GenBank accession number AA047006.

4. The method according to claim 3 wherein the marker gene is hu-*sui1*.

5. The method of claim 1 wherein the step of assessing the expression of said one or more marker genes comprises contacting said sample with one or more probes which detect mRNA which is differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells.

6. The method of claim 5 wherein the mRNA detected hybridizes under high stringency conditions to a nucleic acid probe having a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10.

7. The method of claim 1 wherein the step of assessing the expression of said one or more marker genes comprises detecting marker protein encoded be said one or more marker genes.

8. A method according to claim 7 wherein the one or more marker genes is selected from the group of genes expressing an RNA transcript which hybridizes under high stringency conditions to a nucleic acid probe selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10.

9. The method according to claim 7 wherein the marker gene is selected from the group consisting of

a gene which encodes the polypeptide of SEQ ID NO:27;

a gene which encodes the polypeptide of SEQ ID NO:28;

a gene which encodes the polypeptide of SEQ ID NO:29;

a gene which encodes the polypeptide of SEQ ID NO:29;

hu-sui;

human tubulin-folding cofactor E gene;

human myeloblast KIAA0132 gene; and

the human fetal heart gene, the cDNA of which is identified as GenBank accession number AA047006.

10. The method of claim 7 wherein the one or more marker proteins are detected by contacting the sample with one or more antibodies which bind said marker proteins.

11. A method according to claim 10 wherein the marker gene is *hu-sui1*.

12. A method for diagnosing hepatocellular carcinoma comprising:

obtaining a liver tissue sample from a patient;

assessing the level of expression of one or more marker genes in the sample, which marker genes are differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells, a reduction in the level of expression of said one or more marker genes in the sample as compared to the expression level in noncancerous liver tissue being indicative of hepatocellular carcinoma in the sample.

13. The method according to claim 12 wherein the one or more marker genes is selected from the group of genes expressing an RNA transcript which hybridizes to a nucleic acid probe selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4,

SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, and SEQ ID NO:10.

14. The method according to claim 12 wherein the marker gene is selected from the group consisting of

a gene which encodes the polypeptide of SEQ ID NO:27;
a gene which encodes the polypeptide of SEQ ID NO:28;
a gene which encodes the polypeptide of SEQ ID NO:29;
a gene which encodes the polypeptide of SEQ ID NO:29;
hu-*sui*;
human tubulin-folding cofactor E gene;
human myeloblast KIAA0132 gene; and
the human fetal heart gene, the cDNA of which is identified as GenBank accession number AA047006.

15. The method according to claim 14 wherein the marker gene is hu-*sui1*.

16. A method for identifying alterations in gene expression patterns in a biological sample that are induced by hepatitis B x antigen comprising the steps of

obtaining a biological sample;
contacting said sample with a probe which detects an mRNA which is differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells; and
detecting expression of a gene encoding said mRNA detected by said probe.

17. The method of claim 16 wherein the biological sample is liver tissue.

18. The method of claim 17 wherein the liver tissue comprises a hepatocellular carcinoma.

19. A method for identifying patterns in gene expression in a biological sample that are altered by hepatitis B x antigen comprising the steps of

obtaining a biological sample;

contacting said sample with an antibody which detects a protein which is differentially expressed in HBxAg[+] cells as compared with HBxAg[-] cells; and

detecting expression of a gene encoding said protein detected by said antibody.

20. The method of claim 19 wherein the biological sample is liver tissue.

21. The method of claim 20 wherein the liver tissue comprises a hepatocellular carcinoma.

SEQUENCE LISTING

<110> Feitelson, Mark

<120> Molecular Markers for Diagnosing Hepatocellular Carcinoma

<130> 36seq

<140> Not Yet Assigned
<141> 1999-01-15

<150> 60/072,938
<151> 1998-01-29

<160> 26

<170> PatentIn Ver. 2.0

<210> 1
<211> 488
<212> DNA
<213> Homo sapiens

<400> 1
tgtccccact cttcaaagcc aagatggtag ctgccatccc tggggagcct ggaaccaggc 60
aatgttcggg ggaggcaggg gacaggctgg aacctggtga agtcttaaag taaactcctc 120
ctatcggggt gtagaaggga atctgttaat caaacagagc aatattagaa aggctacaga 180
ggtcaactca gtggaacacg gttctcccaa acagattttg taattccgaa aatccacgca 240
tgcgcaaaca tacgcataca ctcccatgtt cctggacagt ttatagctac cataacctgg 300
cattttccaa aacataccat gtagactctt ggatacacaa ggtaatttta gagccacatt 360
acgatgaacc ttttaaaaag ttatcattta tttttatntt cccccactgg ctgtattata 420
agacaatttt tatatgtgat atgtatttac cttagtgtgt taaataaaca cnggcattcc 480
ctaaaaaa 488

<210> 2
<211> 90
<212> DNA
<213> Homo sapiens

<400> 2
acctgcccgg gcggccgctc gagccctata gtgagtcgta ttagaggccc ttgtagcgtg 60
aagacgacag aaagggcgtg gtgcggaggg 90

<210> 3
<211> 360
<212> DNA
<213> Homo sapiens

<400> 3
tgaaaaccaa actggcggga tggaagcaga ttattctgcc atttttccag gtctttgagt 60
tgcacgtcaa atctggggct gatcacccca cacttgttta gcctgcctgt gaggttcaca 120
acaattttcc cagctctgtg gtcatcaatg atttcaaatt cgccaatgta accatgcttc 180
atcatcacag tgagaaaccg gacgatgact ttggagcacg gcctaataag cacctggcgt 240
ttgcctctct tttcggcatt gttgatactc ttgagagcat ctgccaggac attcatgcgc 300
accattgtgg cggcgcggaa aacctgcccg ggcggccgct cgaaatccat atgactagta 360

<210> 4
<211> 90
<212> DNA
<213> Homo sapiens

<400> 4
acctgttgag gcacttttgt ttcttgggca aaaatacagt ccaatggaga gtatcattgt 60
ttttgtaccg ccctccgcac cacgccctaa 90

<210> 5
<211> 199
<212> DNA
<213> Homo sapiens

<400> 5
tacgggaagg cgaagaaaag aatagagaag atagggaaat tagaagataa aaacatactt 60
ttagaggaaa aaagataaat ttaaacctga aaagtaggaa gcagaagaaa aaagacaagc 120
taggaaacaa aaagctaagg gcaaaatgta ctgaaaatca agatcaagcg agcttttgcc 180
cttctgctcc acgggaggt 199

<210> 6
<211> 658
<212> DNA
<213> Homo sapiens

<400> 6
tcagggaaag ccttcacaga tcagtcaata aatacggtgc catgggatgt gccttgcaca 60
ccacgggcac tcacatcttg aatgctggtc cactggaggc ccttgggttg ggcgggagca 120
aggcctactt ctgcttcctc aggacaactt ccccacctct gtcctgggac cacctgcccg 180
cctgggcctg cagtgactaa ggacgctgct cccactccag gggccagtga cagagagcag 240
ctatacagag ggcccacccc gcaggatcct tgacaggagc tgagacagaa caaactgctg 300
cttgtctccc taccctgggg gctgtgatat tcttggtaac atctctgagc tggtctgtga 360
ggtcacttcc tcttttaaca ctgttgagga gactccaaac cctctgtctt ctgctcgtct 420
tctcatgtcg attgggcacc agccattctc aggcaccaga gcacagcccc acacgggtgc 480
cccatcagac agggctgccc acagcagcct cctacacctg aactgggttt ctctgcacac 540
tcacagccgt ctcaccagct caatgagctg ctggatgttt ttgttttggt tcgacaagcc 600
gttcctgatg ttttccagta ggcatctctt caattcaaat atggcttcac tgtaagcc 658

<210> 7

<212> DNA
<213> Homo sapiens

<400> 7
tcgaggcggt accgggnncc gatttgtagc gtgaanacga cagaaaaggc gtggtgcgga 60
gggcggtgta gcgtgaagac gacagaaagg gcgtggcgga gggcggtacc tacccttccc 120
accctgacgg ggagtgctgc cccgtgtgcc gagactgcaa ctacgaggga aggaaggtgg 180
cgaatggcca ggtgttcacc ttggatgatg aaccctgcac ccggtgcacg tgtcagctgg 240
gaagaggtga gctgtgagaa ggttccctgc cagcgggcct gtgccgaccc tgccctgctt 300
cctggggact gctgctcttc ctgtccagat tccctgtctc ctctggaaga aaagcagggg 360
ctctcccctc acggaaatgt ggcattcagc aaagctggtc ggagcctgca tggaagacac 420
tgaggcccct gtcaactgta gctcctgtcc tgggcccccg acagcatcac cctcgaaggc 480
cggtgcttca tctcctccag ctccttttaa gaacgaactt gatgaaaaca cagactttac 540
ctacaagccc ggcaggagct catggtccac actcactcgc tttggggctg acagccactt 600

<210> 8
<211> 532
<212> DNA
<213> Homo sapiens

<400> 8
aagccccccc ctttcccttn ttttttcttnc ccaannnggg natnccggcc cnttnggntg 60
aaccaantta acnnatcccc naaanatggt nccccnntgg aataanccnt gggttactnt 120
ccaaaccaat tttaccccta tttactttaa atggaattaa ancctccctn tttcattttt 180
aaagggatca nggtgaaaat cccnttgnaa aacccccccna accaaaancc ccttaaantt 240
nattttccct tccccgggaa cttncnaacc cngtaaaaaa anaaanaang gtttccncnn 300
aaatncnttt tttccgcccn ctttatatcg ncgccttngn aaaananaaa aaancccccc 360
nccccccggg tggtttttnt nccccggattt aaaaancccc nntctttttt ccaaaggttt 420
ccggntnccc ccaaaacncc aanacccncn ntcggncccc cctttttttgc cccngntnng 480
ggcccccncc naaaaatctt ntccccccnc cntcatcncn cgctngnnaa tn 532

<210> 9
<211> 425
<212> DNA
<213> Homo sapiens

<400> 9
gctttttttt ttttttttt ttttttttt tttatttctg aaaacaagtt ttatttaaat 60
aagggtttaa tacattacac ataacattaa aactgaaggg gaaaaaaaaa ccaaaaacca 120
gtttgttact tcacatggca ttgggcagct gctgctatta agttgcaagc tctacagcta 180
gctacatgac tgatggatca gtttgagatt tgttcccttg tcaaaagttt aactctgata 240
gaaggttggc ctcacattct gatgtttgga catcccttag ctaggatatg tctggtcgaa 300
cagacctttg tggcaagcca gatgtcctat cacctcccta gcggtaagag ggcctctttg 360
agctctgtcc acctagtcag gttggaaaca ccaggggatc taccacaaaa gctcccttct 420
aatag 425

<210> 10

<212> DNA
<213> Homo sapiens

<400> 10
cagncggtac cggggnnccg attgtagcgt gaaagacgac agaaaaggcg tggtgcggan 60
ggcggttgta ncgtgaanac gacagaaacg ggcgtggtgg agggagtgga ggtacctgcc 120
cgggcggccg ctcgagccct atagtgagtc gtattacccc tatagtgagt cgtattaggc 180
cttgtagcgt gaaaacnaca gaaagggcgt ggtgcggagg gcggtgtacc gtgaanacaa 240
cagaaagggc gtggtgcgga gggcggttgt agcgtgaana cgacagaacc cgcncgagcc 300
ctatagtgag tcgtattang ccctatagtg agtcgtatta ngccctatag tgagtcgtat 360
taggccctat agtgagtcgt attaccccta tagtgagtcg tattagccct atagtgagtc 420
gtattangcc ctatagtgag tcgtattacg ccctataatg agtcgtatta ngccctatag 480
tgagtcgtat tangccctat agtgagtcgt attaggccct atagtgagtc gtattangcc 540
ctatagtgag tcctattagg ccttgtancg tgaaaacaac agccctatan ttgagtcgta 600
ttaaaaatcc atatgactag tnnatcctct agantcnacc tgccaggcat gcaagctttc 660
cctatanttg agtcgtatta aaagcttggc gta 693

<210> 11
<211> 3607
<212> DNA
<213> Homo sapiens

<400> 11
cctcgctgcg ctcatggcgt cgcccgggca ttcagatttg ggagaagtag ccccagaaat 60
aaaagcatca gagagacgaa cagctgtggc cattgcagat ttggaatgga gagaaatgga 120
aggagatgat tgcgagttcc gttatggaga tggtacaaat gaggctcagg acaatgattt 180
tccaacagtg gagagaagca ggcttcaaga aatgctgtca cttttgggcc tagagacgta 240
ccaggtccag aaactcagcc tccaggactc tctgcagatc agttttgaca gtatgaagaa 300
ctgggccccc caggttccca aagacttgcc ctggaatttc ctcaggaagt tgcaggccct 360
caatgctgat gccaggaata ccactatggt gctggacgtg ctcccagacg ccaggcctgt 420
ggagaaggag agccagatgg aagaggagat catctactgg gacccagctg atgaccttgc 480
tgccgacatt tattcctttt ctgagctgcc cacccctgat acgccagtga acccccttaga 540
ccttctctgt gccctgctgc tctcctcaga cagtttcctg caacaagaaa tagcgttgaa 600
aatggccctc tgccagtttg cactcccact cgtgttgcct gactcggaga accactacca 660
tacatttctg ctgtgggcca tgcggggcat tgtgaggaca tggtggtccc agccccccaag 720
gggcatgggg agcttccggg aagacagcgt ggtcttgtcc agggcgcccg ccttcgcctt 780
cgtgcgcatg gacgtcagta gcaactccaa gtcccagctt ctcaacgccg tcctcagccc 840
gggccacagg cagtgggact gcttctggca tcgggacctc aacttgggca ccaatgcccg 900
ggagatttcg gatgggttgg tagaaatttc ctggtttttt cccagcggaa gggaggactt 960
ggacattttc ccagaacctg tggcctttct gaacctgaga ggtgacatcg ggtctcactg 1020
gctgcagttt aagctcttga cagaaatctc ctccgctgtg tttatattga ctgacaatat 1080
cagtaagaag gaatacaaat tgctgtactc catgaaggag tcaaccacaa aatactactt 1140
catcctgagt ccctaccgtg ggaagcgcaa cacaaacctg agatttctga ataagttaat 1200
tcctgtgctg aaaatagacc actcacatgt cctggtaaag gtcagcagca ctgacagcga 1260
cagcttcgtg aagaggatcc gggccatcgt tgggaatgtg ctgcgggcac cctgcaggcg 1320
ggtatctgtg gaggacatgg cgcacgcagc ccgcaaactg ggcctaaagg tcgacgagga 1380
ctgtgaggag tgtcagaaag cgaaagaccg gatggagagg attaccagga aaatcaaaga 1440

ctcggatgcc tacagaaggg acgagctgag gctgcagggg gacccctgga gaaaggcagc 1500
ccaagtggag aaggagttct gccagctcca gtgggccgtg gacccccctg agaagcacag 1560
ggctgagctg aggcggcggc tgctagaact tcgaatgcag cagaacggcc atgattcctc 1620
ctcgggggtg caggagttca tctcggggat cagcagcccc tccttgagtg agaagcagta 1680
cttcctgagg tggatggagt ggggcctggc acgggtggcc cagccgcgac tgagacagcc 1740
tccggagacg cttctcaccc tgagaccaaa gcacgggggc accacagacg tgggggagcc 1800
gctctggcct gagcccctag gggtggaaca cttcttgcgg gagatgggac agttttatga 1860
ggctgagagc tgtcttgtgg aggcagggag gctgccggca ggccagaggc gttttgccca 1920
cttcccaggc ttggcctcgg agctgctgct gacagggctg cctctggagc taatcgatgg 1980
gagcacgctg agcatgcccg tccgctgggt cacagggctc ctgaaggagc tgcacgtccg 2040
actggagaga cggtcaaggc tggtggttct gtcaaccgtc ggggtgccag gcacgggcaa 2100
gtccacactc ctcaacacca tgtttgggct gcggtttgcc acagggaaga gctgcggtcc 2160
tcgaggggcc ttcatgcagc tcatcacagt ggctgagggc ttcagccagg acctgggctg 2220
tgaccacatc ctggtgatag actccggggg cttgataggt ggggccttga cgtcagctgg 2280
ggacagattt gagctggagg cttccttggc cactctgctc atgggactga gcaatgtcac 2340
cgtgatcagt ctagctgaaa ccaaggacat tccagcagct attctgcatg catttctgag 2400
gttagaaaaa acggggcaca tgcccaacta ccagtttgta taccagaacc ttcatgatgt 2460
atctgttccc ggccctaggc ccagagacaa gagacagctc ctggatccac ctggtgacct 2520
gagcagggct gcagcccaga tggagaaaca gggcgacggc ttccgggcac tggcaggcct 2580
ggccttctgc gaccctgaga agcagcacat ctggcacatc ccaggcctgt ggcacggagc 2640
acctcccatg gccgcagtga gcttggccta cagtgaagcc atatttgaat tgaagagatg 2700
cctactcgaa aacatcagga acggcttgtc gaaccaaaac aaaaacatcc agcagctcat 2760
tgagctggtg agacggctgt gagtgtgcag agaaacccag ttcaggtgta ggaggctgct 2820
gtgggcagcc ctgtctgatg gggcacccgt gtggggctgt gctctggtgc ctgagaatgg 2880
ctggtgccca atcgacatga gaagacgagc agaagacaga gggtttggag tctcctcaac 2940
agtgttaaaa gaggaagtga cctcacagac cagctcagag atgttaccaa gaatatcaca 3000
gcccccaggg tagggagaca agcagcagtt tgttctgtct cagctcctgt caaggatcct 3060
gcggggtggg ccctctgtat agctgctctc tgtcactggc ccctggagtg ggagcagcgt 3120
ccttagtcac tgcaggccca ggcgggcagg tggtcccagg acagaggtgg ggaagttgtc 3180
ctgaggaagc agaagtaggc cttgctcccg cccaacccaa gggcctccag tggaccagca 3240
ttcaagatgt gagtgcccgt ggtgtgcaag gcactcccat ggcaccgtat ttattgactg 3300
atctgtgaag gcttccctga cccctgccca ggaagagttc actggtcgct ctgttgtgcc 3360
ccacagcact ttgttatacc tctgccacac acttcacgca gcgcgttgta actcatgtgt 3420
ttacatgtct gtcccccccag actgtgagct ccttgagggc agggactgta cattctccag 3480
ctctgtgtcc ccagggcctg gcacattgta gacgcttaat aaatgtctgt taaatgaaaa 3540
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3600
aaaaaag 3607

<210> 12
<211> 753
<212> DNA
<213> Homo sapiens

<400> 12
cttaccgaca gacagacgct gggacccacg acgacagaag gcgccgatgg ccgcctgctg 60
agccctgcgc ggggcagggg gtctggaacc agacagagcc tgaacctgcc gccaccagcc 120
tgctgagcct gtgcttcctg agaacagcag gggtctgggt acccccccatg tacctctggg 180
tccttggtcc catctacctc ctcttcatcc accaccatgg ccggggctac ctccggatgt 240

ccccactctt caaagccaag atggtagctg ccatccctgg gagcctggaa ccaggcaatg 300
ttcgggggag gcagggaca ggctggaacc tggtgaagtc ttaaagtaga ctcctcctat 360
cggggtgtag aagggaatct gttaatcaaa cagagcaata ttagaaaggc tacagaggtc 420
aactcagtgg aacatggttc tcccaaacag attttgtaat tccgaaaatc cacgcatgcg 480
caaacatacg catacactcc catgttcctg gacagtttat agctaccata acctggcatt 540
ttccaaaaca taccatgtag actcttggat acacaaggta attttagggc cacattagga 600
tgaacctttt aaaaggttat gcatttattt ttatgtttcc ccactagctg tattatagga 660
caatttttat atgtgatatg tatttacctt agtgtgttaa ataaacactg gcatttcaaa 720
aaaaaaaaaa aaaaaaaaaa aaagcggccg ctg 753

<210> 13
<211> 2251
<212> DNA
<213> Homo sapiens

<400> 13
gaattctagg gggaacctca gacccccctc actccttcag ggggaggtga tgggacccct 60
tcctcaccca ggggccctga gtccccccga ctggcagcag ggccctctcc ctgctggcac 120
ctgggagcca tgcatgaatc aaggagtcgc tggacagagc ctgggtgttc ccagtgctgg 180
tgcgaggacg ggaaggtgac ctgtgaaaag gtgaggtgtg aagctgcttg ttcccaccca 240
attccctcca gagatggtgg gtgctgccca tcgtgcacag gctgttttca cagtggtgtc 300
gtccgagctg aaggggatgt gttttcacct cccaatgaga actgcaccgt ctgtgtctgt 360
ctggctggaa acgtgtcctg catctctcct gagtgtcctt ctggcccctg tcagaccccc 420
ccacagacgg attgctgtac ttgtgttcca gtgagatgct atttccacgg ccggtggtac 480
gcagacgggg ctgtgttcag tgggggtggt gacgagtgta ccacctgtgt ttgccagaat 540
ggggaggtgg agtgctcctt catgccctgc cctgagctgg cctgcccccg agaagagtgg 600
cggctgggcc ctgggcagtg ttgcttcacc tgccaggagc ccacaccctc gacaggctgc 660
tctcttgacg acaacggggt tgagtttccg attggacaga tctggtcgcc tggtgacccc 720
tgtgagttat gcatctgcca ggcagatggc tcggtgagct gcaagaggac agactgtgtg 780
gactcctgcc ctcacccgat ccggatccct ggacagtgct gcccagactg ttcagcaggc 840
tgcacctaca caggcagaat cttctataac aacgagacct tcccgtctgt gctggaccca 900
tgtctgagct gcatctgcct gctgggctca gtggcctgtt cccccgtgga ctgccccatc 960
acctgtacct acccctttcca ccctgacggg gagtgctgcc ccgtgtgccg agactgcaac 1020
tacgagggaa ggaaggtggc gaatggccag gtgttcacct tggatgatga accctgcacc 1080
cggtgcacgt gccagctggg agaggtgagc tgtgagaagg ttccctgcca gcgggcctgt 1140
gccgaccctg ccctgcttcc tggggactgc tgctcttcct gtccagattg ccctgtctcc 1200
tctggaagaa aagcaggggc tctcccctca cggaaaatgt ggcattcagc aaagctggtc 1260
gggagcctgc atggagacac tgaggcccct gtcaactgta gctcctgtcc tgggcccccg 1320
acagcatcac cctcgaggcc ggtgcttcat ctcctccagc tcctttaag aacgaacttg 1380
atgaaaacac agactttacc tacaagcccg gcaggagctc atggtccaca ctcactcgct 1440
ttggggctga cagccacttt cccaggggag cctggggcct cccctcgact ctcaccaggg 1500
ccttcgaccc ctccaggagc ccccactcta cctctagctt ccccaggggc tcctcagcca 1560
cctcctgtga ctccagagcg ctcgttctca gcctctgggg cccagatagt gtccaggtgg 1620
cctcctctgc ctggcaccct cctgacggaa gcttcagcac tttccatgat ggaccccagc 1680
ccctcgaaga cccccatcac cctcctcggg cctcgcgtgc tttctcccac cacctctaga 1740
ctctccacag cccttgcagc caccacccac cctggccccc agcagccccc agtgggggct 1800
tctcgggggg aagagtccac catgtaagga ggtcactgtg tccgggagac tctggagaga 1860
ggacctctgc cagtggccca gggtgtgtgc agggcacctc caaggatgaa cctggtgggg 1920

atgcctgggc tccctcctgc acgggccctg gtgaggatgg aagaccccca ggctggatg 1980
taaccttgtt cccaagaagt gtttggaatg tgctgtaaga atggaggaag tcgtttccac 2040
tgtcagcatc ctcccctgga ccgcgtggct ggctcatctt ttgagaaggg ttgggactgc 2100
caagttttcc tggaggaaga gttgcgtccg gctgggattc cactcactgg gactgtaccg 2160
ccaggtgtca tgcgtctttc tgaggtttcc tgattaaagg ttgtttcggt ttcctaaaaa 2220
aaaaaaaaaa aaaaaaaaaa aaaaagcggc c 2251

<210> 14
<211> 541
<212> DNA
<213> Homo sapiens

<400> 14
gccacaatgg tgcgcatgaa tgtcctggca gatgctctca agagtatcaa caatgccgaa 60
aagagaggca aacgccaggt gcttattagg ccgtgctcca aagtcatcgt ccggtttctc 120
actgtgatga tgaagcatgg ttacattggc gaatttgaaa tcattgatga ccacagagct 180
gggaaaattg ttgtgaacct cacaggcagg ctaaacaagt gtggggtgat aagccccaga 240
tttgacgtgc aactcaaaga cctggaaaaa tggcagaata atctgcttcc atcccgccag 300
tttggtttca ttgtactgac aacctcagct ggcatcatgg accatgaaga agcaagacga 360
aaacacacag gagggaaaat cctgggattc tttttctagg gatgtaatac atatatttac 420
aaataaaatg cctcatggac tctggtgctt ccaaaaaaaa aaaaaaaaaa aaaaaaaaaa 480
aaaaaaaaaa aaaaaaaaaa aaaaattaaa aaaaaaaaaa aaaaaaaaaa aaaaagcggc 540
c 541

<210> 15
<211> 660
<212> DNA
<213> Homo sapiens

<400> 15
gccgccgycg aggattcagc agcctccccc ttgagccccc tcgcttcccg acgttccgtt 60
ccccccctgcc cgccttctcc cgccaccgcc gccgccgcct tccgcagccg tttccaccga 120
ggaaaaggaa tcgtatcgta tgtccgctat ccagaacctc cactctttcg acccctttgc 180
tgatgcaagt aagggtgatg acctgcttcc tgctggcact gaggattata tccatataag 240
aattcaacag agaaacggca ggaagaccct tactactgtc caagggatcg ctgatgatta 300
cgataaaaag aaactagtga aggcgtttaa gaaaaagttt gcctgcaatg gtactgtaat 360
tgagcatccg gaatatggag aagtaattca gctacagggt gaccaacgca agaacatatg 420
ccagttcctc gtagagattg gactggctaa ggacgatcag ctgaaggttc atgggtttta 480
agtgcttgtg gctcactgaa gcttaagtga ggatttcctt gcaatgagta gaatttccct 540
tctctccctt gtcacaggtt taaaaacctc acagcttgta taatgtaacc atttggggtc 600
cgcttttaac ttggactagt gtaactcctt catgcaataa actgaaaaga gccatgcaaa 660

<210> 16
<211> 18
<212> PRT
<213> Homo sapiens

<400> 16

<212> PRT
<213> Homo sapiens

<400> 20
Cys Gly Val Trp Asn Gln Thr Glu Pro Glu Pro Ala Ala Thr Ser
1 5 10 15

<210> 21
<211> 16
<212> PRT
<213> Homo sapiens

<400> 21
His His His Gly Arg Gly Tyr Leu Arg Met Ser Pro Leu Phe Lys Cys
1 5 10 15

<210> 22
<211> 16
<212> PRT
<213> Homo sapiens

<400> 22
Pro Cys Pro Glu Leu Ala Cys Pro Arg Glu Glu Trp Arg Leu Gly Pro
1 5 10 15

<210> 23
<211> 24
<212> PRT
<213> Homo sapiens

<400> 23
Asp Pro Ser Arg Ser Pro His Ser Thr Ser Ser Phe Pro Arg Gly Ser
1 5 10 15

Ser Ala Thr Ser Cys Asp Ser Arg
20

<210> 24
<211> 21
<212> PRT
<213> Homo sapiens

<400> 24
His Pro Pro Asp Gly Ser Phe Ser Thr Phe His Asp Gly Pro Gln Pro

1 5 10 15

Leu Glu Asp Pro Cys
20

<210> 25
<211> 13
<212> PRT
<213> Homo sapiens

<400> 25
Lys Ser Ile Asn Asn Ala Glu Lys Arg Gly Lys Arg Cys
1 5 10

<210> 26
<211> 14
<212> PRT
<213> Homo sapiens

<400> 26
Asp His Glu Glu Arg Arg Arg Lys His Thr Gly Gly Lys Cys
1 5 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/01894

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :G01N 33/53; C12Q 1/68; C07H 21/04
US CL : 435/6, 7.1; 536/24.31
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 7.1; 536/24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,534,405 A (MORIARITY et al) 09 July 1996, see entire document.	1-21
A	FEITELSON et al. Hepatitis B virus X antigen in the pathogenesis of chronic infections and the development of hepatocellular carcinoma. Am. J. of Pathology. April 1997, Vol. 150, No. 4, pages 1141-1157, see entire document.	1-21

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search
11 APRIL 1999

Date of mailing of the international search report
19 MAY 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Authorized officer
Jehanne Sousaya

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/01894

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 2,3,4,6,8,9,13,14,15
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

because the claims are directed to probes and polypeptides with SEQ ID NOS that were not available for examination. Errors were detected in Applicants sequence disk, rendering the sequences unsearchable.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/01894

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN, Caplus, Medline, Biosis
search terms: Hepatocellular carcinoma, HBxAg-, HBxAg+, mRNA, antibody, liver, human-sui1, human tubulin-folding cofator E gene, human myeloblast KIAA0132 gene, and human fetal heart gene